Change point detection by Filtered Derivative with p-Value : Choice of the extra-parameters

Pierre, Raphael Bertrand, Doha Hadouni

Change point detection by Filtered Derivative with p-Value : Choice of the extra-parameters

Pierre Raphaël BERTRAND*†,
Doha HADOUNI*†

December 9, 2015

Abstract: This paper deals with off-line change point detection using the FDpV method. The Filtered Derivative with p-Value method (FDpV) is a two-step procedure for change point analysis. In the first step, we use the Filtered Derivative (FD) to select a set of potential change points, using its extra-parameters - namely the threshold for detection, and the sliding window size. In the second one, we calculate the p-value for each change point in order to only retain the true positives (true change points) and discard the false positives (false alarms). We give a way to estimate the optimal extra-parameters of the function FD, in order to have the fewest possible false positives and non-detected change points (ND). Indeed, the estimated potential change points may differ slightly from the theoretically correct ones. After setting the extra-parameters, we need to know whether the absence of detection or the false alarm has more impact on the Mean Integrated Square Error (MISE), which prompts us to calculate the MISE in both cases. Finally, we simulate some examples with a Monte-Carlo method to better understand the positive and negative ways the parametrisation can affect the results.

Keywords: Change points detection; Filtered Derivative with p-Value; Filtered Derivative extra-parameters; Mean Integrated Square Error (MISE); Impact on MISE.

Introduction

Change point detection is an important problem in various applications: signal processing [13], global warming [35], magnetospheric dynamics [36], neuro-physiological studies [33, 22, 21], motion of chemical or physical particles [29], finance [11, 37], health [25]... Most of the previous examples concern detection of change on the mean of series derived from the original one, as the series of energy calculated by the wavelet analysis [25] and the series of Hurst index [11, 36, 37]. However, in all those cases, we still detect change on the mean of the derived series, that is change on the mean value of the Hurst series [11, 36, 37] or change of the mean value of the wavelet transform for the series of energy [25]. To sum up, change point detection on the mean is a relevant question in many applications.
On the other hand, in statistics, the change point analysis field has been studied for more than forty years [17, 4, 15] or [23, 20, 32] for an updated overview. Depending on the method of data acquisition, we distinguish two kinds of change point detection :

*Laboratoire de Mathématiques, UMR CNRS 6620 and Université Clermont Auvergne, France.
†Research supported by grant ANR-12-BS01-0016-01 entitled *"Do Well B."*

- We observe the whole time series and we want to detect all the change point *a posteriori* or *offline*, see e.g. [8, 14].
- We observe the time series and we want to detect a change point as soon as possible. It is the *online* change point detection, see e.g [18, 9].

In this work, we only consider the *'a posteriori'* detection which is called change point analysis in the statistical literature. We describe our framework with a toy model in section 1.
At the beginning of 21st century, the method used for this kind of problem was the Penalized Least Square Criterion. This algorithm is based on the minimisation of the contrast function when the number of change points is known [3]. When the number of change point is unknown, many authors use the penalized version of the contrast function [27]. From a computational point of view, the PLS method uses dynamic programming algorithms and requires matrix operations. Therefore, the time and memory complexity of PLS algorithm is of order $\mathcal{O}(n^2)$, where n denotes the size of the dataset. Due to the data deluge, the sizes of datasets become larger and larger, to the point where the computational complexity of this statistical method has become a challenge, see e.g [24] for internet traffic, [3, 10, 11, 19] for economics or High Frequency finance, [25, 2] for heartbeat series and health.
Among the different methods for *a posteriori* change detection, the use of a Filtered Derivative function has been introduced by [6, 4]. The advantage of the Filtered Derivative method is the time and memory complexity, both of order $\mathcal{O}(n)$ [8, 31, 34, 30]. On the other hand, Filtered Derivative method leads to many false discoveries of change points. Recently, a new method called Filtered Derivative with p-value (FDpV) has been introduced [8]. FDpV is a two-step procedure: the first step is based on the Filtered Derivative function and detects the potential change points. In the second step we calculate their p-value to eliminate the false alarms. In [30], the first step is still based on Filtered Derivative, but the second step consists on increasing the window size A in order to find the true positives.
Yet, the problem of the false discoveries with the Filtered Derivative function, in the first step, was not resolved, even if in the second step, the number of false alarms and non-detections drop substantially. Indeed, most of the false discoveries at the first step will be discarded during the second step by calculating their p-values. However, these calculations still increase the computational time. This problem led us to think of a way to minimise the number of false alarms and also the number of undetected change points in the first step. Furthermore, we investigate the impact of the false positives and the undetected change points on the Mean Integrated Square Error (MISE).
The rest of this article is organized as follows. In Section 1, we describe the problem of change point analysis with a toy model and we give some comparison criterion. In Section 2, we recall the method of the Penalized Least Square and the Filtered derivative with p-Value in order to analyse the problem of change points. In section 3, we expand on the Filtered Derivative with p-Value method by providing a method to choose the extra parameters of Step 1 and we show which impact is more important on MISE. All the technical proofs are postponed in appendices.

1 Change point analysis

In this section, we describe the problem of the change point analysis in a toy model that will be used throughout the sequel of this work. Then, we give some comparison criterion.

1.1 Toy model

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a series indexed by the time $\mathbf{t} = 1, 2, \ldots, n$. We assume that a segmentation $\tau = (\tau_1, \ldots, \tau_K)$ exists such that :

- X_t is a family of independent identically distributed (iid) random variables for $t \in (\tau_k, \tau_{k+1}]$,
- $k = 0, \ldots, K$, where by convention $\tau_0 = 0$ and $\tau_{K+1} = n$.

The most simple model is $\mathbf{X} \sim \mathcal{N}(\mu(\cdot), \sigma^2)$ a sequence of independent standard Gaussian variables such that $X_t \in \mathcal{N}(\mu(t), 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian law with mean μ and variance σ^2. The function of time $t \longmapsto \mu(t)$ is piecewise constant that is to say $\mu(t) = \mu_k$ for all $t \in (\tau_k, \tau_{k+1}]$, see eg. Fig. 3 and Fig. 4. To sum up, we have :

- a configuration of K change points $\tau = (\tau_1, \ldots, \tau_K)$ enlarged, by convention, by adding $\tau_0 = 0$ and $\tau_{K+1} = n$,
- associated to the configuration of mean values $\mu = (\mu_0, \ldots, \mu_K)$,
- $X_t \in \mathcal{N}(\mu_k, \sigma)$, for $t \in (\tau_k, \tau_{k+1}]$ and for all $k = 0, \ldots, K$.
- For notational convenience, we define the configuration of shifts $\delta = (\delta_1, \ldots, \delta_K)$ where $\delta_k = \mu_k - \mu_{k-1}$, for $k = 1, \ldots, K$.
- The minimal distance between two consecutive change points is defined by

$$L_0 = \inf\{|\tau_{k+1} - \tau_k|, for\ k = 0, \ldots, K\}.$$

- The minimal absolute value of the shifts is

$$\delta_0 = \inf\{|\delta_k|,\ k = 1, \ldots, K\}. \tag{1.1}$$

Let us also recall the definition of the cumulative distribution function for standard Gaussian law

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-u^2}{2}} du \qquad \text{and} \qquad \Psi(x) = 1 - \Phi(x). \tag{1.2}$$

1.2 The Comparison Criterion

We have to estimate the configuration $\tau = (\tau_1, \ldots, \tau_K)$ and the values of the mean $\mu = (\mu_0, \mu_1, \ldots, \mu_K)$. We denote the corresponding estimates by $\hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_{\widehat{K}})$ and $\hat{\mu} = (\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_{\widehat{K}})$. Stress that in real life situations the number of change points is also unknown and is estimated by $\widehat{K}$. In this frame, the comparison criterion concerning the different methods for change point analysis are :

1. The quality of estimation. For one sample, this quality can be measured by :
   - The number of estimated change points. More precisely, the absolute value of the difference between the number of estimated and the number of true change points and $|\widehat{K} - K|$.

- **The accuracy of the estimation of the change point**. It is the distance between the true change points and the estimated change points, as defined by

$$d^2(\tau, \widehat{\tau}) \;\; = \;\; \sum_{k=1}^{K} |\tau_k - \widehat{\tau}_{j(k)}|^2, \tag{1.3}$$

  where $\widehat{\tau}_{j(k)}$, for each $k = 1, \ldots, K$, denotes the potential change point which is the closer to the right change point τ_k.

- **The integrated square error (ISE)**. Actually, we can reformulate the problem as a problem of estimation of a noisy signal, see *eg.* [1, 12]. The signal is

$$s(t) = \sum_{k=0}^{K} \mu_k \times \mathbf{1}_{(\tau_k, \tau_{k+1}]}(t)$$

  where we have set by convention $\tau_0 = 0$ and $\tau_{K+1} = n$. Therefore the estimated signal is

$$\widehat{s}(t) = \sum_{k=0}^{\hat{K}} \widehat{\mu}_k \times \mathbf{1}_{(\widehat{\tau}_k, \widehat{\tau}_{k+1}]}(t)$$

  and the integrated square error (ISE) is defined by

$$ISE = \sum_{t=1}^{n} (\widehat{s}(t) - s(t))^2$$

2. **The mean value of estimations:** a result obtained for just one simulation can be hazardous. So, we have to do M simulations, with e.g. $M = 1,000$. Then, we calculate the mean integrated square error (MISE) and the histogram of $\widehat{K}$ with the percentage of the true changes, or the mean and standard deviation of the misestimation of the number of change point : $(\widehat{K} - K)$.

3. The **time complexity** and the **memory complexity**: it is the mean CPU (Central Processing Unit) time for estimating $\widehat{s}$ and the amount of memory is used.

2 Some methods for change point analysis

In this section, we expose two methods for change point analysis. The first one is the Penalized Least Square (PLS, see Subsection 2.1) and the second one is the Filtered Derivative with p-Value (FDpV, see Subsection 2.2). Before going further, let us give some notations. We will denote by

$$\widehat{\mu}(X, [u, v]) := \frac{1}{(v - u + 1)} \times \sum_{t=u}^{v} X_t \tag{2.1}$$

the empirical mean of the variables X_t calculated on the box $t \in [u, v]$. Stress that we can also define $\widehat{\mu}(X, Box)$ for an open box or a semi-open interval.

2.1 Penalized Least Square method (PLS)

Set $\mathcal{S}_{\mathcal{K}} = \{\tau = (\tau_1, \ldots, \tau_K) \text{ such that } card(\tau) = K\}$, where $card(\tau)$ denotes the dimension of the change points configuration τ.

1^{st} **case** : the number of change points K is **known**

For each configuration of change $\tau \in \mathcal{S}_{\mathcal{K}}$, we can define

$$\widehat{\mu}_k = mean(X, [\tau_k + 1, \tau_{k+1}]) \ for \ k = 0, \ldots, K \tag{2.2}$$

where $mean(X, Box)$ denotes the mean of the family X_t for the indices $t \in Box$ as defined by (2.1). Next, we search the configuration of change points $\widehat{\tau}_K \in \mathcal{S}_{\mathcal{K}}$ which minimises the square error $\mathrm{Q}(\tau)$ defined by

$$\mathrm{Q}(\tau) = \sum_{k=0}^{K} \sum_{t=\tau_k+1}^{\tau_{k+1}} |X_t - \widehat{\mu}_k|^2. \tag{2.3}$$

That is to say such that

$$\widehat{\tau}_K = \underset{\tau_K \in \mathcal{S}_{\mathcal{K}}}{\operatorname{argmin}} \mathrm{Q}(\tau).$$

2^{nd} **case** : the number of change points K is **unknown**

We remark that minimising the function $\mathrm{Q}(\tau)$ with an unknown number of changes will lead to consider the trivial configuration of changes $\tau^\star = (1, 2, \ldots, n)$ as optimal. To avoid this drawback, we add a penalty term proportional to the length of the change point configuration. Eventually, we want to minimise

$$pen(K) = \mathrm{Q}(\widehat{\tau}_K) + \beta \times K \quad \text{for } K = 0, \ldots, n.$$

The parameter β adjusts the trade-off between minimising the square error $\mathrm{Q}(\widehat{\tau}_K)$ and minimising the dimension of the change point configuration $card(\widehat{\tau}^K) = K$. Indeed, a large value of β would allow to detect only the most significant change points, while a low value of β produces a high number of changes, with many false detections. Thus, different choices of the penalty coefficient β are possible. According to the criterion of information AIC and the Schwarz criterion, [27] suggests to use a positive β_n which converges to 0 when the series size n converges to infinity. For this model, [38] has proved the consistency of the Schwarz criterion, with

$$\beta_n = \frac{2\sigma^2(\ln n)}{n}.$$

The same choice is proposed in [27, 28]. In [12], the proposed choice is

$$\beta_n = \frac{\sigma^2}{n} \times \left[2 + 5 \times \ln(\frac{n}{K})\right]$$

where σ^2 is the variance assumed to be constant and known and n the size of the series. In Fig. 1 below, we have plotted the contrast function and the penalized contrast function [27, 28].

We clearly see in the figure 2.1 that the penalized contrast is almost horizontal. Thus, the minimal value fluctuates largely depending on the choice of the parameter β.

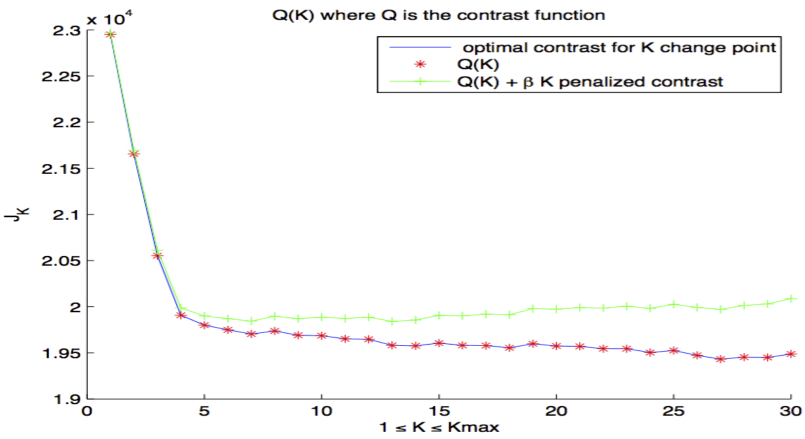


Figure 1: *Blue with red crosses : the contrast function* $\mathrm{Q}(\widehat{\tau}^K)$*; green : the penalized contrasted* $pen(K)$.

2.2 Filtered Derivative with p-Value method (FDpV)

In this subsection, we describe the FDpV method which is based on two procedures. In the first procedure, we use the function of Filtered Derivative to select a set of potential change points, whereas in the second one, we calculate the p-value for each change point in order to keep only the true positive. Precisely, the method is defined as follows:

Step 1 : Filtered Derivative

The first step (FD selection) depends on two parameters: the window size A and the threshold C_1.

1. **Computation of the Filtered Derivative function :**
The Filtered Derivative function is defined as the difference between the estimators of the mean computed in two sliding windows respectively to the right and to the left of the time t, both of size A, with the following formula :

$$\begin{aligned} FD(t,A) &= \widehat{\mu}(X,[t+1,t+A]) - \widehat{\mu}(X,[t-A+1,t]), \\ & \quad for\ A < t < n-A, \end{aligned} \tag{2.4}$$

where $mean(X, Box)$ denotes the mean of the family X_t for the indices $t \in Box$ as defined by (2.1). This method consists on filtering data by computing the estimators of the parameter μ before applying a discrete derivation. So this construction explains the name of the algorithm, so called Filtered Derivative method [6, 4]. Next, remark that quantities $A \times FD(t, A)$ can be inductively calculated by using

$$A \times FD(t+1,A) = A \times FD(t,A) + X(t+1+A) - 2X(t+1) + X(t-A+1) \tag{2.5}$$

Thus, the computation of the whole function $t \longmapsto FD(t)$ for $t \in [A, n-A]$ requires $\mathcal{O}(n)$ operations and the storage of n real numbers.

2. **Determination of the potential change points**
Let us point that the absolute value of the Filtered Derivative $|FD|$ presents hats at the vicinity of the change points as seen on the figure 2 below.

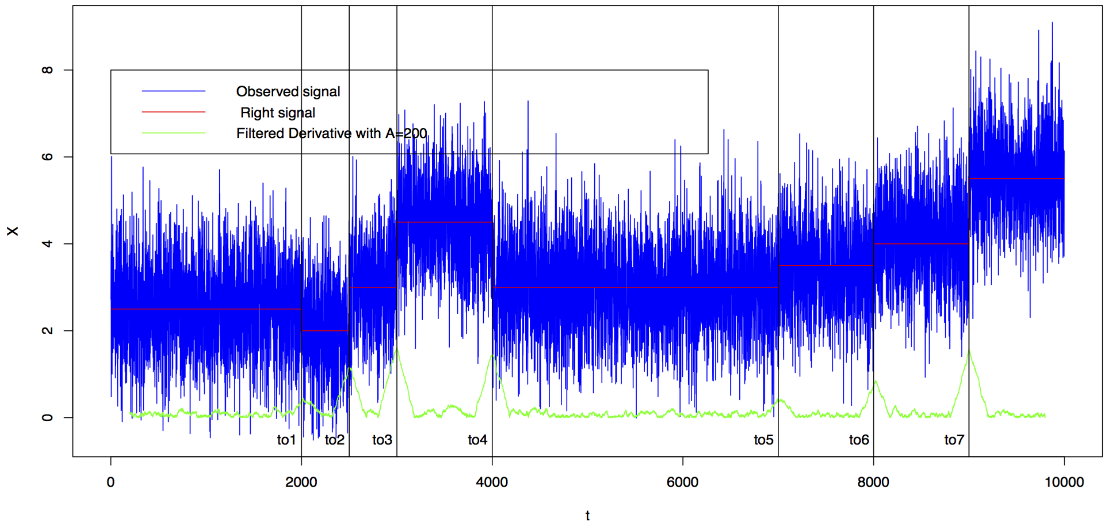


Figure 2: *A graph of the Filtered Derivative function*

Potential change points τ_k^*, for $k = 1, \dots, K^*$, are selected as local maxima of the absolute value of the filtered derivative $|FD(t, A)|$ where moreover $|FD(\tau_k^*, A)|$ exceed a given threshold C_1. When there is a signal without noise ($\sigma = 0$), we get spikes of width $2A$ and height $|\mu_{k+1} - \mu_k|$ at each change point τ_k as we can see in the figure 3 below. For this reason, we select as first potential change point τ_k^* the global maximum

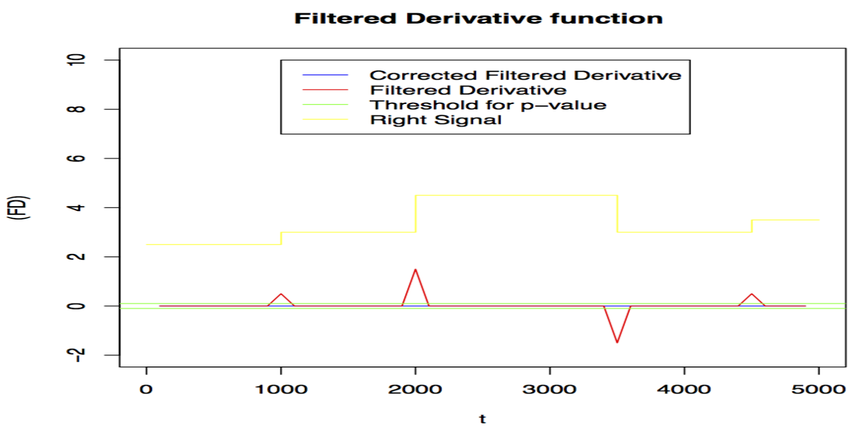


Figure 3: *Filtered Derivative function without noise ($\sigma = 0$).*

of the function $|FD_k(t, A)|$, then we define the function FD_{k+1} by putting to 0 a vicinity of width $2A$ of the point τ_k^* and we iterate this algorithm while $|FD_k(\tau_k^*, A)| > C_1$, see [8]. When there is noise (e.g. $\sigma = 1$), we get the following landscape, see Fig. 4 below.

Step 2 : p-Value

1. **Elimination of false alarm**
   A potential change point τ_k^* can be an estimator of a true change point or a false alarm.

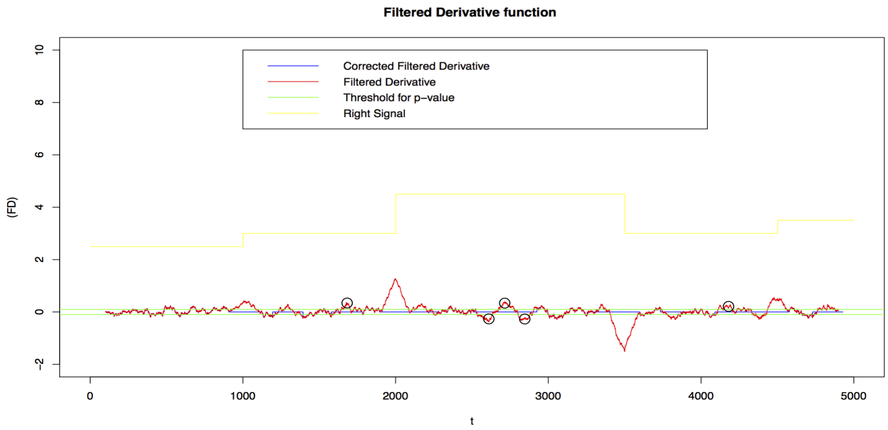


Figure 4: *Filtered Derivative function with noise ($\sigma = 1$). The blue circles correspond to false alarms.*

In the case of a true change point, an error of estimation on the location of the change exists. Thus, we have to cancel a small vicinity of size ε_k around each point τ_k^*, see [7, 8]. Then, for each segment, we calculate an estimation of the mean

$$\widehat{\mu}_k := \widehat{\mu}(X, [\tau_k + \varepsilon_k, \tau_{k+1} - \varepsilon_{k+1}]), \tag{2.6}$$

where $\widehat{\mu}(X, Box)$ is defined by (2.1), and, as in [7], $\varepsilon_k = \left\lceil 5 \times \left(\frac{\sigma}{\delta_k}\right)^2 \right\rceil$ where $\lceil x \rceil$ denotes the ceiling function of the real number x.

Remark 2.1 *From definition (1.1), δ_0 is the lower bound of the shifts, thus we can deduce the following upper bound*

$$\varepsilon_k \quad \leq \quad \varepsilon_0 \; := \; \left\lceil 5 \times \left(\frac{\sigma}{\delta_0}\right)^2 \right\rceil. \tag{2.7}$$

The standard deviation σ can be empirically estimated. Next, we can use this bound in Formula (2.6) and set

$$\widehat{\mu}_k := \widehat{\mu}(X, [\tau_k + \varepsilon_0, \tau_{k+1} - \varepsilon_0]). \tag{2.8}$$

After that, we eliminate the false detections in order to keep (as possible) only the true change points. In [7], we apply the following hypothesis testing, for all $1 \leq k \leq K$:

$$(H_{0,k}) : \widehat{\mu}_k = \widehat{\mu}_{k+1} \quad \text{ versus } \quad (H_{1,k}) : \widehat{\mu}_k \neq \widehat{\mu}_{k+1}$$

where the terms $\widehat{\mu}_k$ are defined by (2.8). By using this second single hypothesis test, we calculate the p-values $p_1^*, \ldots, p_K^*$ associated to each potential change point $\tau_1^\star, \ldots, \tau_{K^\star}^\star$.

2. **p-value computation**
We choose the statistic Student T. Indeed, under the null hypothesis, $t_k^\star$ has a Student distribution of degree $d = N_k + N_{k-1} - 2$ such that $N_k = \left\{(\tau_0^* - \varepsilon_0) - (\tau_k^* + \varepsilon_0)\right\}$, where

$$t_k^* = \frac{\widehat{\mu}_k - \widehat{\mu}_{k-1}}{\sqrt{\frac{S_{k-1}^2}{N_{k-1}} + \frac{S_k^2}{N_k}}}, \tag{2.9}$$

and the sample standard deviation is :

$$S_k = \sqrt{\left(\frac{1}{N_k}\sum_{t=\tau_k+\varepsilon_0}^{\tau_{k+1}-\varepsilon_0} X_t^2\right) - \widehat{\mu}_k^2}, \tag{2.10}$$

By construction, $d > 2A - 4\varepsilon_0$, thus for $A > 30$ the distribution of t_k^* is approximatively Gaussian and we can set

$$p_k^\star = 2 \times \left\{1 - St_d(|t_k|)\right\} \simeq 2 \times \left\{1 - \Phi(|t_k|)\right\} \tag{2.11}$$

where St_d denotes the cumulative distribution function of a Student law of degree d and Φ the cumulative distribution function of the zero mean standard Gaussian law as given by (1.2). In [8], we only keep the change points corresponding to a p-value smaller than a fixed threshold p_2^*. Consequently, Step 2 is much more selective and allows us to deduce an estimator of the piecewise constant map $t \longmapsto \mu(t)$.

3 How to choose the extra-parameters of Step 1 (Filtered Derivative)?

All the change points methods depend on extra-parameters which must be well chosen. The filtered derivative method depends on two extra-parameters, namely the window size A and the threshold C_1.

3.1 Criterion for choosing the extra-parameters of FD

In the case of *"At Most one Change point "* (AMOC), the usual criteria are :

- Error of type I which corresponds to the Probability of False Alarm (PFA).
- Error of type II which corresponds to the Probability of Non Detection (PND).

see *eg.* [16, 26]. The PND is well suited for AMOC. But for detecting more than one change point, we have to impose that one detected change point has to be at the vicinity of each real change point. Following [7], for each real change point τ_k, we define the local PND as

$$PND_{local}(\tau_k) = \mathbb{P}\left(B_k\right),$$

where

$$B_k = \left\{\forall t \in [\tau_k - A, \tau_k + A], |FD(t, A)| < C_1\right\}.$$

With these notations, we can define the global PND by

$$PND_{global} = \mathbb{P}\Big(\bigcup_{k=1}^{K} B_k\Big). \tag{3.1}$$

Next, we can define the Number of Non Detection (NND) in the vicinity of the right change points as

$$NND(\omega) \;=\; \sum_{k=1}^{K} \mathbf{1}_{B_k}(\omega). \tag{3.2}$$

Furthermore, in case of no change points, the probability of false alarm is defined as :

$$\alpha(A, C_1) \;=\; \mathbb{P}\Big(\tau(C_1, A) \leq n - A\Big),$$

where $\tau(C_1, A)$ is the first hitting time of C_1, that is

$$\tau(C_1, A) := \inf\{t \geq A \; ; \; |FD(t, A)| \geq C_1\}. \tag{3.3}$$

However, type I error is the probability of at least one false alarm and thus appears as a rough criterion see [7]. Following [5], the number of false alarm (NFA) is a more relevant criterion.

Definition 3.1 *The number of false alarms is defined as follows*

$$NFA \;=\; \hat{K} - K + NND. \tag{3.4}$$

3.2 Impact of non detection and false alarm on the MISE

i) From (3.4), we get $|\hat{K} - K| = |NFA - NND|$. Thus non detection and false alarm impact in the same way the criterion $|\hat{K} - K|$.

ii) Clearly, false alarm does not impact the quantity $d^2(\tau, \hat{\tau})$ defined by (1.3), whereas non detection at the right place increases the quantity $d^2(\tau, \hat{\tau})$.

iii) The impacts of false alarm and non detection on MISE are mainly described by the two following propositions (3.3 and 3.4). Stress that, the potential change points can be different from the true change points.

Definition 3.2 (MISE) *Let τ_1, τ_2 be two change times, set*

$$s(t) = \sum_{k=0}^{K} \mu_k \times \mathbf{1}_{(\tau_k, \tau_{k+1}]}(t)$$

the true signal, and denote by

$$\widehat{s}(t) = \sum_{k=0}^{\widehat{K}} \widehat{\mu}_k \times \mathbf{1}_{(\widehat{\tau}_k, \widehat{\tau}_{k+1}]}(t)$$

the estimated signal.
The mean integrated square error between the times τ_1 and τ_2 is then defined by

$$MISE(\tau_1, \tau_2) := \mathbb{E}\left(\sum_{t=\tau_1}^{\tau_2} |\widehat{s}(t) - s(t)|^2\right).$$

Proposition 3.3 (False alarm) *Let τ_1, τ_2 be two successive change points and $\widehat{\tau}_1$ and $\widehat{\tau}_2$ the potential change points such that $\hat{\tau}_1 = \tau_1 + \varepsilon_1$, $\hat{\tau}_2 = \tau_2 - \varepsilon_2$. Furthermore, we assume that $\|\varepsilon_i\|_\infty \leq M_{\varepsilon_i}$, with $i \in \{1, 2\}$ where M_{ε_i} is a finite constant.*

i) Without false alarm : Assume that $\hat{\tau}_1$, $\hat{\tau}_2$ are the two successive potential change points obtained after Step 1. Then

$$MISE(\tau_1,\tau_2) \quad = \quad \sigma^2 + r_{3.5} \tag{3.5}$$

whith

$$\begin{aligned} |r_{3.5}| \quad \leq \quad & M_{\varepsilon_1}\left[\frac{\sigma^2}{\tau_1-\tau_0}+\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2\right]+\dots \\ & \dots + M_{\varepsilon_2}\left[\frac{\sigma^2}{\tau_3-\hat{\tau}_2}+\left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}\right)^2(\mu_1-\mu_2)^2\right]. \end{aligned}$$

ii) With false alarm : Assume that $\hat{\tau}_1$, $\hat{\tau}_2$, and $\hat{\tau}_3$ are the three successive potential change points obtained after Step 1 such as $\hat{\tau}_3$ is the false alarm, and that $\tau_1 < \hat{\tau}_1 < \hat{\tau}_3 < \hat{\tau}_2 < \tau_2$. Then

$$MISE(\tau_1,\tau_2) = 2\sigma^2 + r_{3.6} \tag{3.6}$$

with :

$$\begin{aligned} |r_{3.6}| \quad \leq \quad & M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1-\tau_0}+\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2\right]+\dots \\ & \dots + M_{\varepsilon_2}\left[\frac{\sigma^2}{\hat{\tau}_2-\tau_4}+\left(\frac{\tau_2-\tau_4}{\hat{\tau}_2-\tau_4}\right)^2(\mu_1-\mu_3)^2\right]. \end{aligned}$$

Proof. See Appendix A. □

Proposition 3.4 (Non detection) *Let τ_1, τ_2, and τ_3 be three successive right change points Furthermore, we assume that $\|\varepsilon_i\|_\infty \leq M_{\varepsilon_i}$, with $i \in \{1,2,3\}$ where M_{ε_i} is a finite constant and the quantity ε_i are precisely defined below.*

i) Without non detection : Assume that $\hat{\tau}_1$, $\hat{\tau}_2$, $\hat{\tau}_3$ are the three successive potential change points obtained after Step 1 such that $\hat{\tau}_1 = \tau_1 + \varepsilon_1$, $\hat{\tau}_2 = \tau_2 - \varepsilon_2$, $\hat{\tau}_3 = \tau_3 - \varepsilon_3$ and $\tau_1 < \hat{\tau}_1 < \tau_2 < \hat{\tau}_2 < \hat{\tau}_3 < \tau_3$.

$$MISE(\tau_1,\tau_3) = \sigma^2 + r_{3.7} \tag{3.7}$$

with :

$$\begin{aligned} |r_{3.7}| \quad \leq \quad & M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1-\tau_0}+\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2\right]+\dots \\ & \dots + M_{\varepsilon_2}\left[\frac{\sigma^2}{\hat{\tau}_3-\hat{\tau}_2}+\left(\frac{\hat{\tau}_3-\tau_2}{\hat{\tau}_3-\hat{\tau}_2}\right)^2(\mu_1-\mu_2)^2\right]+\dots \\ & \dots + M_{\varepsilon_3}\left[\frac{\sigma^2}{\tau_4-\hat{\tau}_3}+\left(\frac{\tau_4-\tau_3}{\tau_4-\hat{\tau}_3}\right)^2(\mu_1-\mu_4)^2\right]. \end{aligned}$$

ii) With non detection : Assume that τ_2 is the undetected change point. For notational convenience, let us still denote by $\hat{\tau}_1$ and $\hat{\tau}_3$ the two successive potential change points obtained after Step 1 such that $\hat{\tau}_1 = \tau_1 + \varepsilon_1$, $\hat{\tau}_3 = \tau_3 - \varepsilon_3$ and $\tau_1 < \hat{\tau}_1 < \tau_2 < \hat{\tau}_3 < \tau_3$. Then

$$MISE(\tau_1, \tau_3) = 2\sigma^2 + r_{3.8} \tag{3.8}$$

with :

$$\begin{aligned} |r_{3.8}| \ \leq \ & M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2 (\mu_1 - \mu_0)^2\right] + \dots \\ & \dots + M_{\varepsilon_3}\left[\frac{\sigma^2}{\tau_4 - \hat{\tau}_3} + \left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}\right)^2 (\mu_1 - \mu_3)^2\right]. \end{aligned}$$

Proof. See Appendix B. □

In order to measure which kind of alarm (error) impacts more the MISE, we study the difference between the MISE in the case of the non detection and the MISE in the case of the false alarm. To begin with, we restricted ourselves to the particular case where $r_{1,i} \longrightarrow 0$ with $i \in \{1, 2, 3, 4\}$, which means that $\varepsilon_j = 0$ with $j \in \{1, 2, 3\}$ and the potential change points are the true change points. Moreover, we obtain the same results at the first order in the general case when we put $\varepsilon_i \neq 0$.

Corollary 3.5 (False alarm) *i) Let $\hat{\tau}_1$, $\hat{\tau}_2$ be the two successive potential change points obtained after Step 1. Then*

$$MISE(\tau_1, \tau_2) = \sigma^2$$

ii) Let $\hat{\tau}_1$, $\hat{\tau}_2$, and $\hat{\tau}_3$ be the three successive potential change points obtained after Step 1 such that $\tau_1 = \hat{\tau}_1 < \hat{\tau}_3 < \hat{\tau}_2 = \tau_2$. Then

$$MISE(\tau_1, \tau_2) = 2\sigma^2.$$

iii) The difference between the MISE in the case with false alarm and the case without false alarm is as follows :

$$\begin{aligned} \Delta MISE_{FA} \ &:= \ MISE_{withFA} - MISE_{withoutFA} \\ &= \ \sigma^2 \end{aligned} \tag{3.9}$$

Corollary 3.6 (Non detection) *i) Let $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}_3$ be the three successive potential change points obtained after Step 1 such that $\hat{\tau}_1 = \tau_1$, $\hat{\tau}_2 = \tau_2$ and $\hat{\tau}_3 = \tau_3$. Then*

$$MISE(\tau_1, \tau_3) = 2\sigma^2$$

ii) Next, assume that τ_2 is the undetected change point. For mathematical convenience, let us still denote by $\hat{\tau}_1$ and $\hat{\tau}_3$ the two successive potential change points obtained after Step 1 such that $\tau_1 = \hat{\tau}_1 < \tau_2 < \hat{\tau}_3 = \tau_3$. Then

$$MISE(\tau_1, \tau_3) = \sigma^2 + \frac{(\tau_2 - \tau_1)(\tau_3 - \tau_2)}{\tau_3 - \tau_1}(\delta\mu)^2$$

with $\delta\mu = (\mu_2 - \mu_1)$.

iii) The difference between the MISE in the case with non detection and the case without non detection is as follows :

$$\begin{aligned} \Delta MISE_{ND} &:= MISE_{withND} - MISE_{withoutND} \\ &= \frac{(\tau_2-\tau_1)(\tau_3-\tau_2)}{\tau_3-\tau_1}(\delta\mu)^2 - \sigma^2 \end{aligned} \tag{3.10}$$

Proposition 3.7 *Let us assume that Propositions 3.3 and 3.4 are satisfied. Then*

- *if* $\left(\frac{\delta\mu}{\sigma}\right)^2 > 2$, *then* $\Delta MISE_{ND} > \Delta MISE_{FA}$, *namely the impact of the non detection is more important than the impact of the false alarm .*
- *On the other side, if* $\left(\frac{\delta\mu}{\sigma}\right)^2 < \frac{8}{\tau_3-\tau_1}$, *then* $\Delta MISE_{ND} < \Delta MISE_{FA}$, *namely the impact of the false alarm is more important than the impact of the non detection.*

Proof. We want to show that :

$$\left(\frac{\delta\mu}{\sigma}\right)^2 > 2 \Longrightarrow \Delta MISE_{ND} > \Delta MISE_{FA} \tag{3.11}$$

We have :

$$\Delta MISE_{ND} - \Delta MISE_{FA} = \sigma^2\left[\left(\frac{(\delta\mu)}{\sigma}\right)^2 \frac{(\tau_2-\tau_1)(\tau_3-\tau_2)}{\tau_3-\tau_1} - 2\right]$$

Thus,

$$\begin{aligned} \Delta MISE_{ND} - \Delta MISE_{FA} > 0 &\Longrightarrow \Delta MISE_{ND} > \Delta MISE_{FA} \\ &\Longrightarrow \frac{\delta\mu}{\sigma} > \sqrt{2\frac{\tau_3-\tau_1}{(\tau_2-\tau_1)(\tau_3-\tau_2)}} \end{aligned}$$

Set $L = \tau_3 - \tau_1$, $\lambda = \tau_2 - \tau_1$ where $L > 0$ and $\lambda \in [1, L-1]$
Thus, we have : $f(\lambda) = \sqrt{\frac{2L}{\lambda(L-\lambda)}}$

The minimum of $f(\lambda)$ is reached for $\lambda = L/2$, and the maximum is reached for $\lambda = 1$.
We deduce then :

$$\sqrt{\frac{8}{L}} \leq f(\lambda) \leq \sqrt{\frac{2L}{L-1}} < \sqrt{2}$$

Therefore, $\frac{\delta\mu}{\sigma} > \sqrt{2}$ implies $\Delta MISE_{ND} > \Delta MISE_{FA}$.
On the other hand, when $\frac{\delta\mu}{\sigma} > \sqrt{\frac{8}{L}}$ we have $\Delta MISE_{ND} < \Delta MISE_{FA}$. $\square$

3.3 Bound on the error of type II for Filtered Derivative

Proposition 3.8 *Let $\tau = (\tau_1, \tau_2, \dots, \tau_K)$ be a configuration of K change points, with means $\mu = (\mu_0, \dots, \mu_K)$ and shifts $\delta = (\delta_1, \dots, \delta_K)$ as described in Subsection 1.1. Then*

$$PND_{global} \leq K \times \beta^*(C_1, A),$$

where PND_{global} is defined by (3.1),

$$\beta^*(C_1, A) := \Psi\left(\frac{\delta_0 - C_1}{\sigma}\sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta_0/3}{\sigma}\sqrt{\frac{A}{2}}\right)^2, \tag{3.12}$$

δ_0 is defined by (1.1), and Φ and Ψ are given by (1.2).

Proof. Actually, the proposition Prop. 3.4 is a corollary of [7, Prop 3.2, p 222]. But the [7, Prop 3.2, p 222] deals with the unilateral case whereas the Prop. 3.4 concerns the bilateral case. Firstly, let us point that when $\delta_k < 0$, we can multiply the function FD by (-1), which brings us back to the case $\delta_k > 0$. Therefore, without any restriction, we can assume that $\delta_k > 0$. Secondly, we have $B_k \subset \widetilde{B}_k$, where

$$\widetilde{B}_k = \Big\{\forall t \in [\tau_k - A, \tau_k + A], FD(t, A) < C_1\Big\}.$$

Thus, $\mathbb{P}\big(B_k\big) \leq \widetilde{\mathbb{P}}\big(B_k\big)$. On the other hand, following [7, Prop. 3.2, p 222], we have,

$$\mathbb{P}\left(B_k\right) \leq \Psi\left(\frac{|\delta_k| - C_1}{\sigma}\sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - |\delta_k|/3}{\sigma}\sqrt{\frac{A}{2}}\right)^2. \tag{3.13}$$

Next, by remarking that the right side of (3.13) is a decreasing function of $|\delta_k|$ and recalling that $\delta_0 = \inf_{k=1,\dots,K} |\delta_k|$, we can deduce that

$$\mathbb{P}\left(B_k\right) \leq \beta^*(C_1, A) := \Psi\left(\frac{\delta_0 - C_1}{\sigma}\sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta_0/3}{\sigma}\sqrt{\frac{A}{2}}\right)^2. \tag{3.14}$$

On the other hand, we obviously have

$$PND_{global} \leq \sum_{k=1}^{K} \mathbb{P}(B_k)$$

which combined with (3.14) provides the bound (3.8). This finishes the proof of Proposition 3.8. □

3.4 Control of the number of false alarms

In this subsection, we want to control not only the Probability of False Alarm (PFA) but also the Number of False Alarms (NFA) (see *Definition* 3.1).
Let us denote by $\widetilde{K}$ the number of change point select in step 1 of the FDpV method (FD), then the number of false alarms is $(\widetilde{K} - K)$.
Moreover, in order to control the number of false alarms, we need to choose the extra-parameters A and C_1.

The choice of parameter A

From the subsection 3.3, we can get the feeling that the larger the window size A is, the smaller type I and type II errors will be. This reasoning holds true as long as

$$2 \times A \quad < \quad L_0 := \inf\{|\tau_{k+1} - \tau_k|, k = 1, \dots, K\}. \tag{3.15}$$

Thus, we have to choose a parameter $A < L_0/2$, even if we do not exactly know the quantity L_0.

The choice of parameter C_1

In [7] (Bertrand, 2000), we have $C_1 < \delta_0$ with $\delta_0 = \inf\{|\delta_k|, k = 1, ..., K\}$ where δ_k is the size of the average of μ_k. With different Monte-Carlo simulations (see the subsection 3.5), we note that the best values of C_1 are between 0.1 and 0.2.

3.5 Monte-Carlo simulation

This Monte-Carlo simulation is done for $M = 1000$. Let $(X_1^j, X_2^j, \dots, X_n^j)$ be a sequence of simulated Gaussian random variables where $n = 5000$ or $n = 50,000$ and $j = 1, \dots, M$, with a variance $\sigma^2 = 1$ and a mean $\mu_t = f(t)$ where f is a piecewise constant function with a specified number of change points at different times τ with different means μ.
On each sample, we apply the FDpV method for different values of the extra-parameters A and C_1. We vary the parameter A between 20 and 220 with a step of 10 while we vary the parameter C_1 between 0.1 and 1 with a step of 0.05.

Example 1

The figures 5 and 6 show respectively the variation of the mean number of non detected change points and the variation of the mean number of the false alarms in function of the extra-parameters A and C_1 where $n = 5000$; $\tau = (1000, 1250, 1500, 2000, 3500, 4000, 4500)$ and $\delta\mu = (-0.5, -1, -1.5, 0.5, 1, 1.5)$ with $\delta_0 = 0.5$.

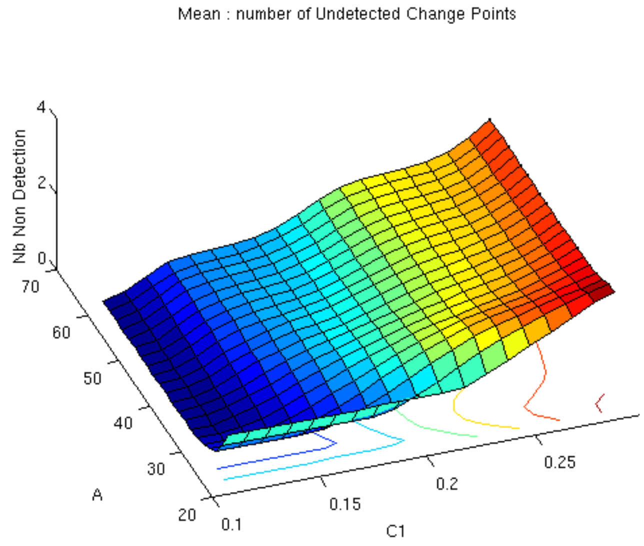


Figure 5: The mean of the number of undetected change points

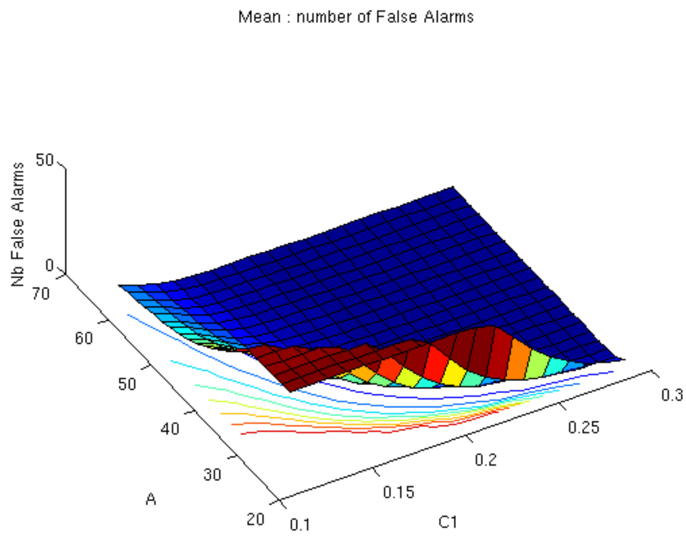
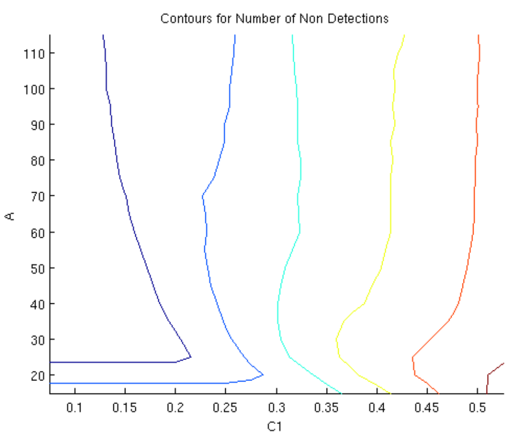
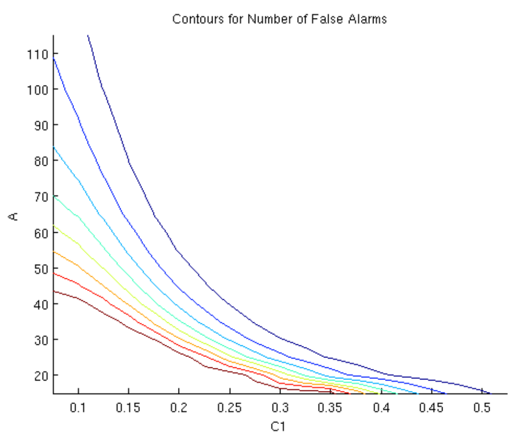


Figure 6: The mean of the number of false alarm

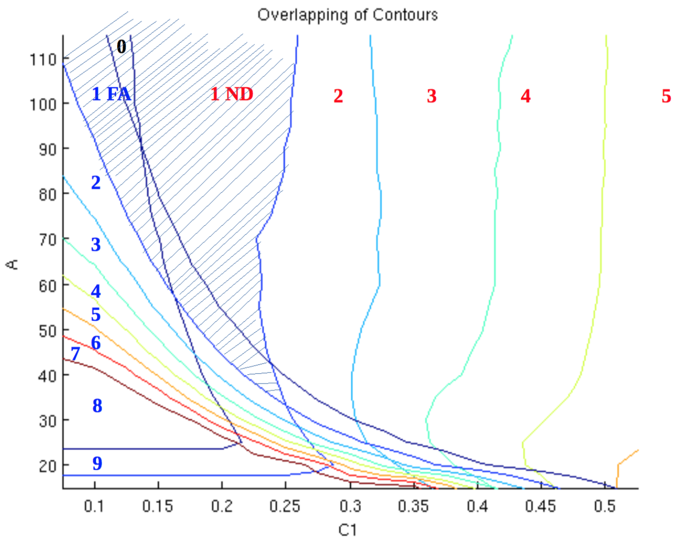
In the figures 7(a) and 7(b) below, we draw the contours of the mean number of undetected change points and those of the mean number of false alarm. These contours help us define an admissible set of extra-parameters. For example, the hatched domain in figure 7(c) corresponds to the intersection of zero or one mean number of non detection and zero or one mean number of false alarm. In this case, a choice of a window $A \geq 70$ and a threshold $C_1 \in [0.15, 0.22]$ insures a mean number of non detection and a mean number of false alarm both smaller than 1.



(a) The contour of the mean number of undetected change points.



(b) The contour of the mean number of false alarm.



(c) The overlap contour of 7(a) and 7(b).

Figure 7: The contours of the figures 5 and 6 and their overlap

Example 2

The figures 8 and 9 shows respectively the variation of the mean number of the undetected change points and the mean number of the false alarms in function of the extra-parameters A and C_1 where $n = 50.000$, $\tau = (10000, 12500, 15000, 20000, 25000, 32500, 35000, 40000, 45000)$ and $\delta\mu = (-2, -1.5, -1, -0.5, 0.5, 1, 1.5, 2)$ with $\delta_0 = 0.5$.

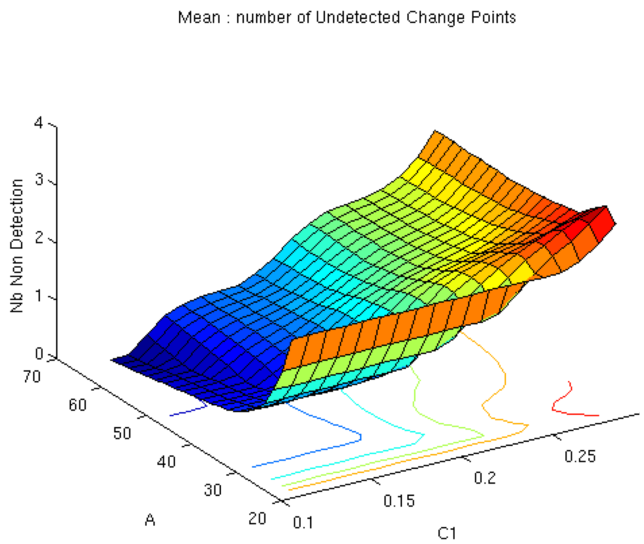


Figure 8: The mean of the number of undetected change points

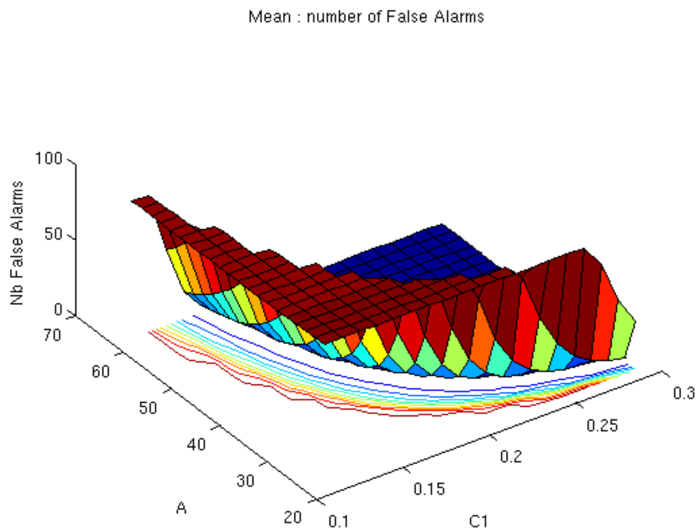
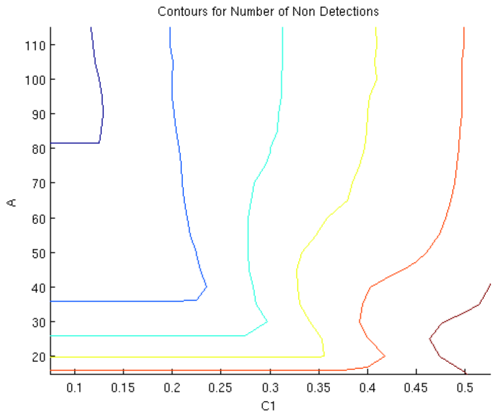
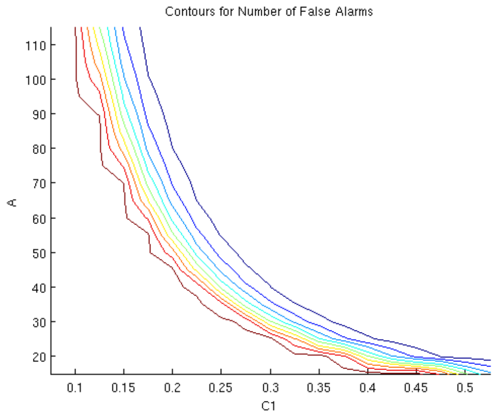


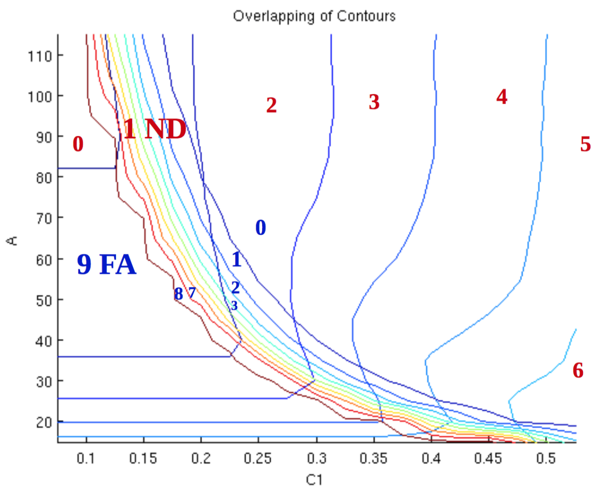
Figure 9: The mean of the number of false alarm



(a) The contour of the number of undetected change points.



(b) The contour of the number of false alarm.



(c) The overlap contour of 10(a) and 10(b).

Figure 10: The contours of the figures 8 and 9 and their overlap

In the figures 10(a) and 10(b) above, we draw the contours of the mean number of undetected change points, those of the mean number of false alarms and their overlap.The numbers on the figure of the overlapping of contours are given as an indication to show the evolution of the mean number of false alarms (blue) and the mean number of undetected change point (red). Let us point that on figure 10(c), the intersection of zero or one non detection and false alarm is void. However, a choice of extra-parameters $A \geq 70$ and a threshold $C_1 \in [0.15, 0.22]$ would insure a mean number of non detection smaller than one and a mean number of false alarm smaller than 8.

Example 3

In this example, we keep the variables n and τ of the example 2 and we take shifts on the mean μ of the change point smaller than those in the example 1 and 2, that is to say $\delta\mu = (-2, -1.5, -1, -0.25, 0.25, 1, 1.5, 2)$ with $\delta_0 = 0.25$. The figures 11 and 12 show the variation of the mean number of the undetected change points and those of the mean number of the false alarms for different values of the extra-parameters A and C_1.

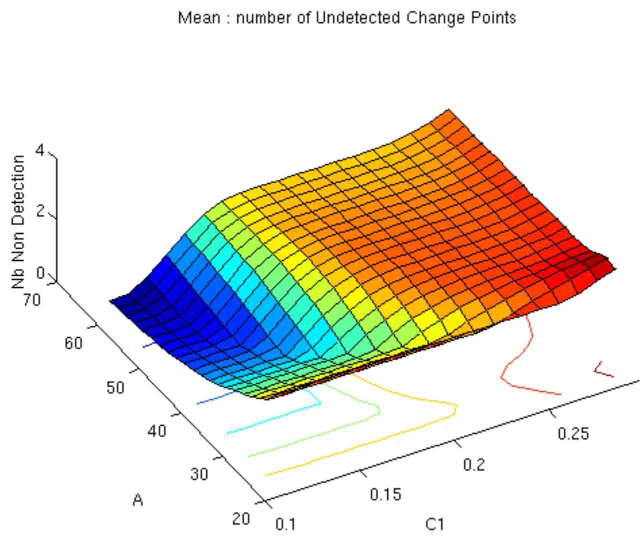


Figure 11: The mean of the number of undetected change points

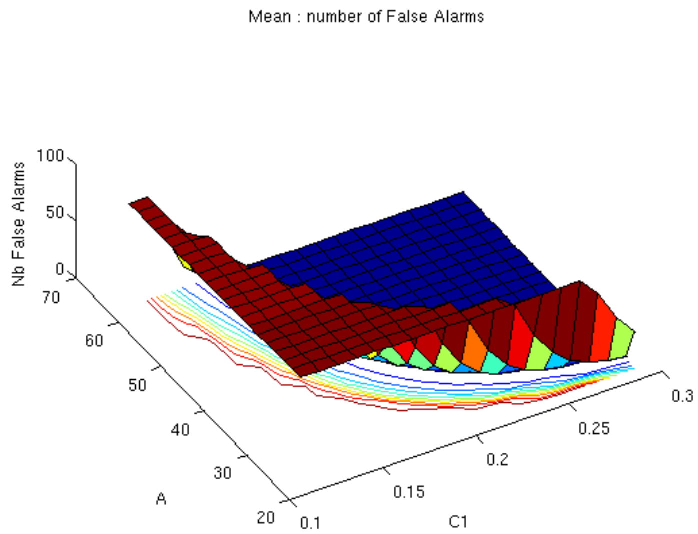
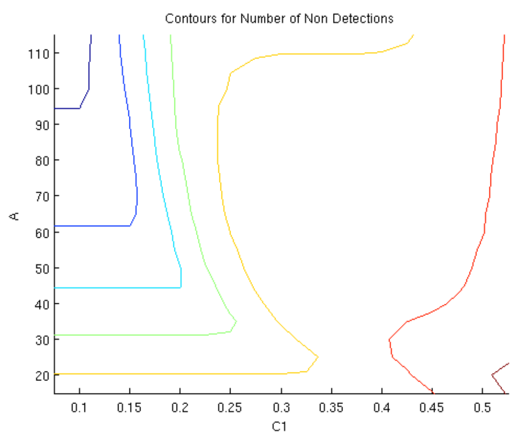
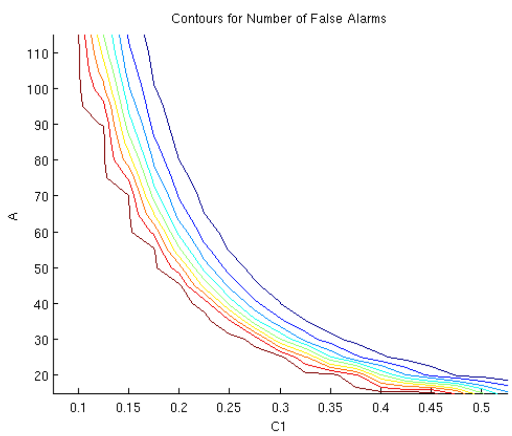


Figure 12: The mean of the number of false alarm

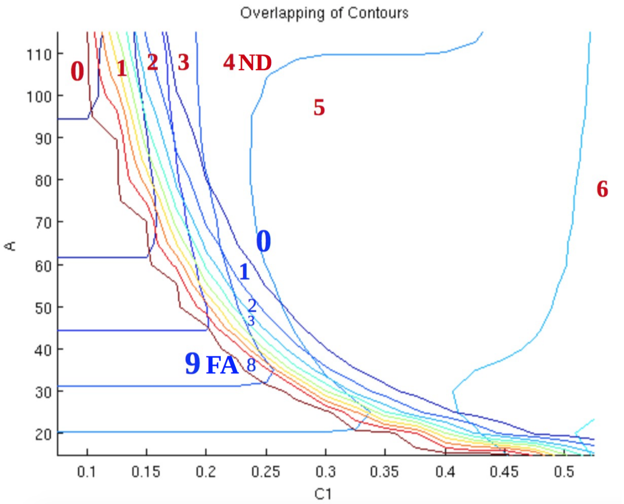
In the figures 13(a) and 13(b) below, we draw the contours of the mean number of undetected change points and those of the mean number of false alarm. A choice of extra-parameters $A \geq 70$ and a threshold $C_1 \in [0.15, 0.22]$ would insure a mean number of non detection smaller than 3 and a mean number of false alarm smaller than 8. But, fixing $C_1 = 0.15$ insures a mean number of non detection smaller than 1 without impact on the mean number of false alarm.



(a) The contour of the number of undetected change points.



(b) The contour of the number of false alarm.



(c) The overlap of the contours 13(a) and 13(b).

Figure 13: The contours of the figures 11 and 12

We note that, in the three examples, when the threshold C_1 is smaller than δ_0 (especially when $C_1 \in [0.15, 0.22]$) and the window size A is large enough without exceeding the value $\frac{L_0}{2}$, we get less undetected change points and less false alarms.

Conclusion

Our analysis suggest a comprehensive way to optimize the Filtered Derivative function extra-parameters in the first step of the FDpV method. By giving the precise values to choose for the window size A and the threshold C_1 where we have to choose $A > 70$ and $C_1 \in [0.15, 0.22]$, we obtain fewer false alarms and undetected change points. Thus, in the second step, we calculate less p-values, gaining computational time and memory.
Another method is used in [30]. The first step is still based on the Filtered Derivative function with the two extra-parameters: the window size A and the threshold C_1. Yet, in the second step, they increase the window size A such that for each value of A they detect a real change point. This algorithm may detect all the true change points. But the problem of the false discoveries is not resolved, and can be enhanced by a good initial choice of the parameters A and C_1 .
On the other hand, by fixing a value for the window size A and the threshold C_1 in Step 1 of the FDpV method, we detect the real change points and, at the same time, we minimise the number of the false discoveries. Furthermore, in Step 2, we eliminate more false positives. Finally, we studied the impact of the false alarms and the undetected change points on the Mean Integrated Square Error in order to determine which one has a more significant impact. To sum up, for the FDpV method, we have specified the values of the window size A and the threshold C_1 in Step 1. In forthcoming studies, we need to find a way to estimate the value of the threshold p_2^* in Step 2 to discard as many false positives as possible.

References

[1] S. Arlot and A. Celisse, *Segmentation of the mean of heteroscedastic data via cross-validation*, Stat. Comput. **21** (2011), no. 4, 613–632. MR 2826696 (2012h:62066)

[2] N. Azzaoui, A. Guillin, F. Dutheil, G. Boudet, A. Chamoux, C. Perrier, J. Schmidt, and P.R. Bertrand, *Classifying heart-rate by change detection and wavelet methods for emergency physicians,*, ESAIM Proc. **45** (September 2014), 48–57.

[3] J. Bai and P. Perron, *Estimating and testing linear models with multiple structural changes*, Econometrica **66** (1998), no. 1, 47–78. MR MR1616121 (98m:62184)

[4] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*, Prentice Hall Information and System Sciences Series, Prentice Hall Inc., Englewood Cliffs, NJ, 1993. MR MR1210954 (95g:62153)

[5] Y. Benjanimi and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. Roy. Stat. Soc **B 57** (1995), 289–300.

[6] A. Benveniste and M. Basseville, *Detection of abrupt changes in signals and dynamical systems: some statistical aspects*, Analysis and optimization of systems, Part 1 (Nice, (1984), Lecture Notes in Control and Inform. Sci., vol. 62, Springer, Berlin, 1984, pp. 145–155. MR MR876686

[7] P. R. Bertrand, *A local method for estimating change points: the "hat-function"*, Statistics **34** (2000), no. 3, 215–235. MR MR1802728 (2001j:62032)

[8] P. R. Bertrand, M. Fhima, and A Guillin, *Off-line detection of multiple change points by the filtered derivative with p-value method*, Sequential Analysis **30 (2)** (2011), 172–207.

[9] P. R. Bertrand and G. Fleury, *Detecting small shift on the mean by finite moving average*, International Journal of Statistics and Management System **3** (2008), 56–73.

[10] P.R. Bertrand, A. Hamdouni, and Khadhraoui S., *Modelling NASDAQ series by sparse multifractional brownian motion.*, Methodology and Computing in Applied Probability **14** (2012), no. 1, 107–124.

[11] S. Bianchi, A. Pantanella, and A. Painese, *Efficient markets and behacioral finance : a comprehensive multifractal model*, Advances in Complex Systems **18** (2015).

[12] L. Birgé and P. Massart, *Minimal penalties for gaussian model selection*, Probab. Theory Related Fields **138** (2007), no. 1-2, 33–73. MR MR2288064 (2008g:62070)

[13] S. Boutoille, *Systèmes de fusion pour la segmentation hors-ligne de signaux gpd multi-porteuses*, Ph.D. thesis, 2007.

[14] S. Boutoille, S. Reboul, and M. Benjelloun, *A hybrid fusion system applied to off-line detection and change-points estimation.*, Information Fusion – Elsevier **11** (October 2010), 325–337.

[15] B. E. Brodsky and B. S. Darkhovsky, *Nonparametric methods in change-point problems*, Mathematics and its Applications, vol. 243, Kluwer Academic Publishers Group, Dordrecht, 1993. MR MR1228205 (95d:62068)

[16] M. Csörgo and L. Horváth, *Limit theorem in change-point analysis*, J. Wiley, New York., 1997.

[17] M. Csörgö and P. Revez, *Strong approximations in probability and statistics*, Akadémiai Kiadö, Budapest, 1981.

[18] P. Fearnhead and Z. Liu, *On-line inference for multiple change point problems.*, Journal of the Royal Statistical Society **69** (2007), 589–605.

[19] M. Frezza, *Modeling the time-changing dependence in stock markets.*, Chaos Solitons and Fractals **45(12)** (2012), 1510–1520.

[20] E. Gombay and D. Serban, *Monitoring parameter change in ar(p) time series models.*, Journal of Multivariate Analysis **100** (2009), 715–725.

[21] S. Grun, M. Diesmann, and A. Aertsen, *Unitary events in multiple single neuron spiking activity. II. non-stationary data.*, Neural Comput. (2002), 81–119.

[22] ———, *Unitary events in multiple single neuron spiking activity: I. detection and significance.*, Neural Comput. **14** (Jan, 2002).

[23] M. Hušková and S. G. Meintanis, *Change point analysis based on the empirical characteristic functions.*, Metrika **63** (2006a), 145–168.

[24] Y. Ji, *Multi-scale internet traffic analysis using piecewise self-similar processes*, IEICE Trans, Commun **E89-B** (August 2006), no. 8, 2125–2133.

[25] N. Khalfa, P.R. Bertrand, G. Boudet, A. Chamoux, and V. Billat, *Heart rate regulation processed through wavelet analysis and change detection.*, Some case studies, Acta Biotheoretica. **60** (2012), 109–129.

[26] A.N. Kolmogorov, Yu.V. Prokhorov, and A.N. Shiryaev, *Probabilist-statistical methods of detecting spontaneously occuring effects*, Proceeding of the Steklov Institute for Mathematics, Moscow (1988).

[27] M. Lavielle and E. Moulines, *Least-squares estimation of an unknown number of shifts in a time series*, J. Time Ser. Anal. **21** (2000), no. 1, 33–59. MR MR1766173 (2001c:62102)

[28] M. Lavielle and G. Teyssière, *Detection of multiple change points in multivariate time series.*, Lithuanian Math. J. **46** (2006), 287–306.

[29] S.C. Lim and L.P. Teo, *Modeling single-file diffusion by step fractional brownian motion and generalized fractional langevin equation.*, Journal of Statistical Mechanics : Theory and Experiment **2009** (2009), no. 8.

[30] M. Messer, M. Kirchener, J. Shiemann, J. Roeper, R. Neininger, and G. Schneider, *A multiple filter test for the detection of rate changes in renewal processes with varying variance*, The Annals of Applied Statistics **8** (2015), no. 4, 2027–2067.

[31] M. Messer and G. Schneider, *The shark function - asymptotoc behavior of the filetered derivative for point process in case of change points*, (2014).

[32] G. Rigaill, T. D. Hocking, F. Bach, and J.P. Vert, *Learning sparse penalties for change-point detection using max margin interval regression.*, Proceedings of the International Conference on Machine Learning (ICML) (2013).

[33] G. Schneider, *Messages of oscillatory correlograms : A spike train model*, Neural Comput **20** (2008), 1211–1238.

[34] Y. S. Soh and V. Chandrasekaran, *High-dimensional change-point estimation: Combining filtering with convex optimization*, Article in Applied and Computational Harmonic Analysis (January 2015).

[35] W. Wang, I. Bobojonov, W.K. Hardle, and M. Odening, *Testing for increasing weather risk*, Stochasitic Environmental Research and Risk Assessment **27** (2013), 1565–1574.

[36] J.A. Wanliss and P. Dobias, *Space storm as phase transition*, Journal of Atmospheric and Solar-Terrestrial Physics **69** (2007), 675–684.

[37] W. Xiao, W. Zhang, and X. Zhang, *Parameter identification for drift fractional brownian motions with application to the chinese stock markets*, Communications in Statistics - Simulation and Computation **44** (2015), 2117–2136.

[38] Y.C. Yao, *Estimating the number of change-points via schwarzs criterion. statistics and probability letters*, J. Nonparametr. Statist. **6** (1998), 181–189.

A Proof of Proposition 3.3

Proof.

i) Without false alarm :

$ISE(\tau_1, \tau_2) = ISE(\tau_1, \hat{\tau}_1) + ISE(\hat{\tau}_1, \hat{\tau}_2) + ISE(\hat{\tau}_2, \tau_2)$
- $ISE(\tau_1, \hat{\tau}_1) = \varepsilon_1(\mu_1 - \hat{\mu_0})^2$
For $\tau_0 = \hat{\tau}_0$ et $\tau_3 = \hat{\tau}_3$, we have :

$\hat{\mu}_0 = \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0 + \frac{(\hat{\tau}_1 - \tau_1)\mu_1 + (\tau_1 - \tau_0)\mu_0}{\hat{\tau}_1 - \tau_0}$

$$\begin{aligned}
\mu_1 - \hat{\mu}_0 &= \mu_1 - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0 - \frac{(\hat{\tau}_1 - \tau_1)\mu_1 - (\tau_1 - \tau_0)\mu_0}{\hat{\tau}_1 - \tau_0} \\
&= \frac{(\hat{\tau}_1 - \tau_0)\mu_1 - (\hat{\tau}_1 - \tau_1)\mu_1 - (\tau_1 - \tau_0)\mu_0}{\hat{\tau}_1 - \tau_0} - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0 \\
&= \frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0
\end{aligned}$$

$$\begin{aligned}
MISE(\tau_1, \hat{\tau}_1) &= \mathbb{E}\left(\varepsilon_1(\mu_1 - \hat{\mu_0})^2\right) \\
&= \mathbb{E}\left[\varepsilon_1\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right] \\
&\leq \|\varepsilon_1\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right]
\end{aligned}$$

$\|\varepsilon_1\|_\infty = sup|\varepsilon_{1_k}| \leq M_{\varepsilon_1}$
$\mathbb{E}\left[\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right] = \frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2(\mu_1 - \mu_0)^2$
Thus :

$$MISE(\tau_1, \hat{\tau}_1) \leq M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2(\mu_1 - \mu_0)^2\right]$$

$ISE(\hat{\tau}_1, \hat{\tau}_2) = (\hat{\tau}_2 - \hat{\tau}_1)(\mu_1 - \hat{\mu_1})^2$
We have : $\hat{\mu}_1 = \mu_1 + \frac{\sigma}{\sqrt{\hat{\tau}_2 - \hat{\tau}_1}} \times U_1 \Rightarrow \mu_1 - \hat{\mu_1} = -\frac{\sigma}{\sqrt{\hat{\tau}_2 - \hat{\tau}_1}} \times U_1$
Thus :

$$\begin{aligned}
MISE(\hat{\tau}_1, \hat{\tau}_2) &= \mathbb{E}\left((\hat{\tau}_2 - \hat{\tau}_1)(\mu_1 - \hat{\mu_1})^2\right) \\
&= (\hat{\tau}_2 - \hat{\tau}_1)\mathbb{E}(\mu_1 - \hat{\mu_1}^2) \\
&= (\hat{\tau}_2 - \hat{\tau}_1)\frac{\sigma^2}{\hat{\tau}_2 - \hat{\tau}_1} \times \mathbb{E}(U_1^2) \\
&= \sigma^2
\end{aligned}$$

$ISE(\hat{\tau}_2 - \tau_2) = \varepsilon_2(\mu_1 - \hat{\mu}_2)^2$
we have : $\hat{\mu}_2 = \frac{\sigma}{\sqrt{\hat{\tau}_3-\hat{\tau}_2}} \times U_2 + \frac{(\tau_2-\hat{\tau}_2)\mu_1+(\hat{\tau}_3-\tau_2)\mu_2}{\hat{\tau}_3-\hat{\tau}_2}$

$$\begin{aligned}
\mu_1 - \hat{\mu}_2 &= \mu_1 - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2 - \frac{(\tau_2-\hat{\tau}_2)\mu_1 - (\tau_3-\tau_2)\mu_2}{\tau_3-\hat{\tau}_2} \\
&= \frac{(\tau_3-\tau_2)\mu_1 - (\tau_2-\hat{\tau}_2)\mu_1 - (\tau_3-\tau_2)\mu_2}{\tau_3-\hat{\tau}_2} - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2 \\
&= \frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}(\mu_1-\mu_2) - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2
\end{aligned}$$

$$\begin{aligned}
MISE(\hat{\tau}_2, \tau_2) &= \mathbb{E}\left(\varepsilon_2(\mu_1-\hat{\mu}_2)^2\right) \\
&= \mathbb{E}\left[\varepsilon_2\left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}(\mu_1-\mu_2) - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2\right)^2\right] \\
&\leq \|\varepsilon_2\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}(\mu_1-\mu_2) - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2\right)^2\right]
\end{aligned}$$

$\|\varepsilon_2\|_\infty = sup|\varepsilon_{2_k}| \leq M_{\varepsilon_2}$
$\mathbb{E}\left[\left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}(\mu_1-\mu_2) - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2\right)^2\right] = \frac{\sigma^2}{\tau_3-\hat{\tau}_2} + \left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}\right)^2(\mu_1-\mu_2)^2$
Thus :

$$MISE(\tau_1, \hat{\tau}_1) \leq M_{\varepsilon_2}\left[\frac{\sigma^2}{\tau_3-\hat{\tau}_2} + \left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}\right)^2(\mu_1-\mu_2)^2\right]$$

Thus, we can deduce that :

$$\begin{aligned}
MISE(\tau_1, \tau_2) &\leq \sigma^2 + \varepsilon_1\left[\frac{\sigma^2}{\hat{\tau}_1-\tau_0} + \left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2\right] \\
&+ \varepsilon_2\left[\frac{\sigma^2}{\tau_3-\hat{\tau}_2} + \left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}\right)^2(\mu_1-\mu_2)^2\right]
\end{aligned} \tag{A.1}$$

ii) With false alarm : $ISE(\tau_1, \tau_2) = ISE(\tau_1, \hat{\tau}_1) + ISE(\hat{\tau}_1, \hat{\tau}_3) + ISE(\hat{\tau}_3, \hat{\tau}_2) + ISE(\hat{\tau}_2, \tau_2)$

$ISE(\tau_1, \hat{\tau}_1) = \varepsilon_1(\mu_1 - \hat{\mu}_0)^2$
we have : $\hat{\mu}_0 = \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}} \times U_0 + \frac{(\hat{\tau}_1-\tau_1)\mu_1+(\tau_1-\tau_0)\mu_0}{\hat{\tau}_1-\tau_0}$

$$\begin{aligned}
\mu_1 - \hat{\mu}_0 &= \mu_1 - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}} \times U_0 - \frac{(\hat{\tau}_1-\tau_1)\mu_1 + (\tau_1-\tau_0)\mu_0}{\hat{\tau}_1-\tau_0} \\
&= \frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1-\mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}} \times U_0
\end{aligned}$$

$$\begin{aligned} MISE(\tau_1, \hat{\tau}_1) &= \mathbb{E}\left(\varepsilon_1(\mu_1 - \hat{\mu}_0)^2\right) \\ &= \mathbb{E}\left[\varepsilon_1\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right] \\ &\leq \|\varepsilon_1\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right] \end{aligned}$$

$\|\varepsilon_1\|_\infty = sup|\varepsilon_{1_k}| \leq M_{\varepsilon_1}$
$\mathbb{E}\left[\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}} \times U_0\right)^2\right] = \frac{\sigma^2}{\hat{\tau}_1-\tau_0} + \left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2 (\mu_1 - \mu_0)^2$
Thus :

$$MISE(\tau_1, \hat{\tau}_1) \leq M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2 (\mu_1 - \mu_0)^2\right]$$

$ISE(\hat{\tau}_1, \hat{\tau}_3) = (\hat{\tau}_3 - \hat{\tau}_1)(\mu_1 - \hat{\mu_1})^2$
We have : $\hat{\mu}_1 = \mu_1 + \frac{\sigma}{\sqrt{\hat{\tau}_3-\hat{\tau}_1}} \times U_1 \Rightarrow \mu_1 - \hat{\mu_1} = -\frac{\sigma}{\sqrt{\hat{\tau}_3-\hat{\tau}_1}} \times U_1$
Thus :

$$\begin{aligned} MISE(\hat{\tau}_1, \hat{\tau}_3) &= \mathbb{E}\left((\hat{\tau}_3 - \hat{\tau}_1)(\mu_1 - \hat{\mu_1})^2\right) \\ &= (\hat{\tau}_3 - \hat{\tau}_1)\mathbb{E}(\mu_1 - \hat{\mu_1})^2 \\ &= (\hat{\tau}_3 - \hat{\tau}_1)\frac{\sigma^2}{\hat{\tau}_3 - \hat{\tau}_1} \times \mathbb{E}(U_1^2) \\ &= \sigma^2 \end{aligned}$$

$ISE(\hat{\tau}_3, \hat{\tau}_2) = (\hat{\tau}_2 - \hat{\tau}_3)(\mu_1 - \hat{\mu_2}^2)$
We have :
$\hat{\mu}_1 = \mu_2 + \frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_3}} \times U_2 \Rightarrow \mu_1 - \hat{\mu_2} = -\frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_3}} \times U_2$
Thus :

$$\begin{aligned} MISE(\hat{\tau}_3, \hat{\tau}_2) &= \mathbb{E}\left((\hat{\tau}_2 - \hat{\tau}_3)(\mu_1 - \hat{\mu_1})^2\right) \\ &= (\hat{\tau}_2 - \hat{\tau}_3)\mathbb{E}(\mu_1 - \hat{\mu_2})^2 \\ &= (\hat{\tau}_2 - \hat{\tau}_3)\frac{\sigma^2}{\hat{\tau}_2 - \hat{\tau}_3} \times \mathbb{E}(U_2^2) \\ &= \sigma^2 \end{aligned}$$

$ISE(\hat{\tau}_2, \tau_2) = \varepsilon_2(\mu_1 - \hat{\mu}_3)^2$
we have : $\hat{\mu}_3 = \frac{\sigma}{\sqrt{\hat{\tau}_2-\tau_4}} \times U_2 + \frac{(\hat{\tau}_2-\tau_2)\mu_1+(\tau_2-\tau_4)\mu_3}{\hat{\tau}_2-\tau_4}$

$$\begin{aligned} \mu_1 - \hat{\mu}_3 &= \mu_1 - \frac{\sigma}{\sqrt{\hat{\tau}_2 - \tau_4}} \times U_2 - \frac{(\hat{\tau}_2 - \tau_2)\mu_1 + (\tau_2 - \tau_4)\mu_3}{\hat{\tau}_2 - \tau_4} \\ &= \frac{\tau_2 - \tau_4}{\hat{\tau}_2 - \tau_4}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\hat{\tau}_2 - \tau_4}} \times U_2 \end{aligned}$$

$$
\begin{aligned}
MISE(\hat{\tau}_2, \tau_2) &= \mathbb{E}\left(\varepsilon_2(\mu_1 - \hat{\mu}_3)^2\right) \\
&= \mathbb{E}\left[\varepsilon_2\left(\frac{\tau_2 - \tau_4}{\hat{\tau}_2 - \tau_4}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\hat{\tau}_2 - \tau_4}} \times U_2\right)^2\right] \\
&\leq \|\varepsilon_2\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_2 - \tau_4}{\hat{\tau}_2 - \tau_4}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\hat{\tau}_2 - \tau_4}} \times U_2\right)^2\right]
\end{aligned}
$$

$\|\varepsilon_2\|_\infty = sup|\varepsilon_{2_k}| \leq M_{\varepsilon_2}$
$\mathbb{E}\left[\left(\frac{\tau_2-\tau_4}{\hat{\tau}_2-\tau_4}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\hat{\tau}_2-\tau_4}} \times U_2\right)^2\right] = \frac{\sigma^2}{\hat{\tau}_2-\tau_4} + \left(\frac{\tau_2-\tau_4}{\hat{\tau}_2-\tau_4}\right)^2 (\mu_1 - \mu_3)^2$
Thus :

$$
MISE(\tau_2, \hat{\tau}_2) \leq M_{\varepsilon_2}\left[\frac{\sigma^2}{\hat{\tau}_2 - \tau_4} + \left(\frac{\tau_2 - \tau_4}{\hat{\tau}_2 - \tau_4}\right)^2 (\mu_1 - \mu_3)^2\right]
$$

Then :

$$
\begin{aligned}
MISE(\tau_1, \tau_2) &\leq 2\sigma^2 + M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2 (\mu_1 - \mu_0)^2\right] \\
&+ M_{\varepsilon_2}\left[\frac{\sigma^2}{\hat{\tau}_2 - \tau_4} + \left(\frac{\tau_2 - \tau_4}{\hat{\tau}_2 - \tau_4}\right)^2 (\mu_1 - \mu_3)^2\right]
\end{aligned} \tag{A.2}
$$

□

B Proof of Proposition 3.4

Proof.

i) Without non detection :
$ISE(\tau_1, \tau_3) = ISE(\tau_1, \hat{\tau}_1) + ISE(\hat{\tau}_1, \hat{\tau}_2) + ISE(\hat{\tau}_2, \tau_2) + ISE(\tau_2, \hat{\tau}_3) + ISE(\hat{\tau}_3, \tau_3)$
- $ISE(\tau_1, \hat{\tau}_1) = \varepsilon_1(\mu_1 - \hat{\mu}_0)^2$
For $\tau_0 = \hat{\tau}_0$ et $\tau_4 = \hat{\tau}_4$, we have :

$\hat{\mu}_0 = \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}} \times U_0 + \frac{(\hat{\tau}_1-\tau_1)\mu_1+(\tau_1-\tau_0)\mu_0}{\hat{\tau}_1-\tau_0}$

$$
\begin{aligned}
\mu_1 - \hat{\mu}_0 &= \mu_1 - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0 - \frac{(\hat{\tau}_1 - \tau_1)\mu_1 - (\tau_1 - \tau_0)\mu_0}{\hat{\tau}_1 - \tau_0} \\
&= \frac{(\hat{\tau}_1 - \tau_0)\mu_1 - (\hat{\tau}_1 - \tau_1)\mu_1 - (\tau_1 - \tau_0)\mu_0}{\hat{\tau}_1 - \tau_0} - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0 \\
&= \frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0
\end{aligned}
$$

$$\begin{aligned}
MISE(\tau_1, \hat{\tau}_1) &= \mathbb{E}\left(\varepsilon_1(\mu_1 - \hat{\mu_0})^2\right) \\
&= \mathbb{E}\left[\varepsilon_1\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right] \\
&\leq \|\varepsilon_1\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1 - \tau_0}} \times U_0\right)^2\right]
\end{aligned}$$

$\|\varepsilon_1\|_\infty = sup|\varepsilon_{1_k}| \leq M_{\varepsilon_1}$
$\mathbb{E}\left[\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1 - \mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}} \times U_0\right)^2\right] = \frac{\sigma^2}{\hat{\tau}_1-\tau_0} + \left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2 (\mu_1 - \mu_0)^2$
Thus :

$$MISE(\tau_1, \hat{\tau}_1) \leq M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2 (\mu_1 - \mu_0)^2\right]$$

$ISE(\hat{\tau}_1, \hat{\tau}_2) = (\hat{\tau}_1, \hat{\tau}_2)(\mu_1 - \hat{\mu_1})^2$
We have :
$\hat{\mu}_1 = \mu_1 + \frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_1}} \times U_1 \Rightarrow \mu_1 - \hat{\mu_1} = -\frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_1}} \times U_1$
Thus :

$$\begin{aligned}
MISE(\hat{\tau}_1, \hat{\tau}_2) &= \mathbb{E}\left((\hat{\tau}_2 - \hat{\tau}_1)(\mu_1 - \hat{\mu_1})^2\right) \\
&= (\hat{\tau}_2 - \hat{\tau}_1)\mathbb{E}(\mu_1 - \hat{\mu_1})^2 \\
&= (\hat{\tau}_2 - \hat{\tau}_1)\frac{\sigma^2}{\hat{\tau}_2 - \hat{\tau}_1} \times \mathbb{E}(U_1^2) \\
&= \sigma^2
\end{aligned}$$

$ISE(\hat{\tau}_2 - \tau_2) = \varepsilon_2(\mu_1 - \hat{\mu}_2)^2$
we have : $\hat{\mu}_2 = \frac{\sigma}{\sqrt{\hat{\tau}_3-\hat{\tau}_2}} \times U_2 + \frac{(\tau_2-\hat{\tau}_2)\mu_1+(\hat{\tau}_3-\tau_2)\mu_2}{\hat{\tau}_3-\hat{\tau}_2}$

$$\begin{aligned}
\mu_1 - \hat{\mu}_2 &= \mu_1 - \frac{\sigma}{\sqrt{\tau_3 - \hat{\tau}_2}} \times U_2 - \frac{(\tau_2 - \hat{\tau}_2)\mu_1 - (\tau_3 - \tau_2)\mu_2}{\tau_3 - \hat{\tau}_2} \\
&= \frac{(\tau_3 - \tau_2)\mu_1 - (\tau_2 - \hat{\tau}_2)\mu_1 - (\tau_3 - \tau_2)\mu_2}{\tau_3 - \hat{\tau}_2} - \frac{\sigma}{\sqrt{\tau_3 - \hat{\tau}_2}} \times U_2 \\
&= \frac{\tau_3 - \tau_2}{\tau_3 - \hat{\tau}_2}(\mu_1 - \mu_2) - \frac{\sigma}{\sqrt{\tau_3 - \hat{\tau}_2}} \times U_2
\end{aligned}$$

$$\begin{aligned}
MISE(\hat{\tau}_2, \tau_2) &= \mathbb{E}\left(\varepsilon_2(\mu_1 - \hat{\mu}_2)^2\right) \\
&= \mathbb{E}\left[\varepsilon_2\left(\frac{\tau_3 - \tau_2}{\tau_3 - \hat{\tau}_2}(\mu_1 - \mu_2) - \frac{\sigma}{\sqrt{\tau_3 - \hat{\tau}_2}} \times U_2\right)^2\right] \\
&\leq \|\varepsilon_2\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_3 - \tau_2}{\tau_3 - \hat{\tau}_2}(\mu_1 - \mu_2) - \frac{\sigma}{\sqrt{\tau_3 - \hat{\tau}_2}} \times U_2\right)^2\right]
\end{aligned}$$

$\|\varepsilon_2\|_\infty = sup|\varepsilon_{2_k}| \leq M_{\varepsilon_2}$
$\mathbb{E}\left[\left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}(\mu_1-\mu_2) - \frac{\sigma}{\sqrt{\tau_3-\hat{\tau}_2}} \times U_2\right)^2\right] = \frac{\sigma^2}{\tau_3-\hat{\tau}_2} + \left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}\right)^2 (\mu_1-\mu_2)^2$
Thus :

$$MISE(\tau_1, \hat{\tau}_1) \leq \varepsilon_2 \left[\frac{\sigma^2}{\tau_3 - \hat{\tau}_2} + \left(\frac{\tau_3 - \tau_2}{\tau_3 - \hat{\tau}_2}\right)^2 (\mu_1 - \mu_2)^2\right]$$

$ISE(\tau_2, \hat{\tau}_3) = (\hat{\tau}_3 - \tau_2)(\mu_1 - \hat{\mu_2})^2$
We have :
$\hat{\mu}_2 = \mu_1 + \frac{\sigma}{\sqrt{\hat{\tau}_3-\tau_2}} \times U_3 \Rightarrow \mu_1 - \hat{\mu}_2 = -\frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_1}} \times U_3$
Thus :

$$\begin{aligned} MISE(\hat{\tau}_1, \hat{\tau}_2) &= \mathbb{E}\left((\hat{\tau}_2 - \hat{\tau}_1)(\mu_1 - \hat{\mu_2})^2\right) \\ &= (\hat{\tau}_2 - \hat{\tau}_1)\mathbb{E}(\mu_1 - \hat{\mu_1}^2) \\ &= (\hat{\tau}_2 - \hat{\tau}_1)\frac{\sigma^2}{\hat{\tau}_2 - \hat{\tau}_1} \times \mathbb{E}(U_3^2) \\ &= \sigma^2 \end{aligned}$$

$ISE(\hat{\tau}_3, \tau_3) = \varepsilon_3(\mu_1 - \hat{\mu_4})^2$
we have : $\hat{\mu}_4 = \frac{\sigma}{\sqrt{\tau_4-\hat{\tau}_3}} \times U_4 + \frac{(\tau_3-\hat{\tau}_3)\mu_1+(\tau_4-\tau_3)\mu_4}{\tau_4-\hat{\tau}_3}$

$$\begin{aligned} \mu_1 - \hat{\mu}_4 &= \mu_1 - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_4 - \frac{(\tau_3 - \hat{\tau}_3)\mu_1 + (\tau_4 - \tau_3)\mu_4}{\tau_4 - \tau_3} \\ &= \frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_4) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_4 \end{aligned}$$

$$\begin{aligned} MISE(\hat{\tau}_3, \tau_3) &= \mathbb{E}\left(\varepsilon_3(\mu_1 - \hat{\mu}_4)^2\right) \\ &= \mathbb{E}\left[\varepsilon_3\left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_4) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_4\right)^2\right] \\ &\leq \|\varepsilon_3\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_4) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_4\right)^2\right] \end{aligned}$$

$\|\varepsilon_3\|_\infty = sup|\varepsilon_{3_k}| \leq M_{\varepsilon_3}$
$\mathbb{E}\left[\left(\frac{\tau_4-\tau_3}{\tau_4-\hat{\tau}_3}(\mu_1-\mu_4) - \frac{\sigma}{\sqrt{\tau_4-\hat{\tau}_3}} \times U_4\right)^2\right] = \frac{\sigma^2}{\tau_4-\hat{\tau}_3} + \left(\frac{\tau_4-\tau_3}{\tau_4-\hat{\tau}_3}\right)^2 (\mu_1-\mu_4)^2$
Thus :

$$MISE(\hat{\tau}_3, \tau_3) \leq M_{\varepsilon_3} \left[\frac{\sigma^2}{\tau_4 - \hat{\tau}_3} + \left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}\right)^2 (\mu_1 - \mu_4)^2\right]$$

Thus, we can deduce that :

$$\begin{aligned} MISE(\tau_1,\tau_3) &\leq 2\sigma^2 + M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1-\tau_0} + \left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2\right] \\ &+ M_{\varepsilon_2}\left[\frac{\sigma^2}{\tau_3-\hat{\tau}_2} + \left(\frac{\tau_3-\tau_2}{\tau_3-\hat{\tau}_2}\right)^2(\mu_1-\mu_2)^2\right] \\ &+ M_{\varepsilon_3}\left[\frac{\sigma^2}{\tau_4-\hat{\tau}_3} + \left(\frac{\tau_4-\tau_3}{\tau_4-\hat{\tau}_3}\right)^2(\mu_1-\mu_4)^2\right] \end{aligned} \tag{B.1}$$

ii) With non detection : $ISE(\tau_1,\tau_3) = ISE(\tau_1,\hat{\tau}_1)+ISE(\hat{\tau}_1,\tau_2)+ISE(\tau_2,\hat{\tau}_3)+ISE(\hat{\tau}_3,\tau_3)$
$ISE(\tau_1,\hat{\tau}_1) = \varepsilon_1(\mu_1-\hat{\mu}_0)^2$
For $\tau_0=\hat{\tau}_0$ et $\tau_4=\hat{\tau}_4$, we have :

$\hat{\mu}_0 = \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0 + \frac{(\hat{\tau}_1-\tau_1)\mu_1+(\tau_1-\tau_0)\mu_0}{\hat{\tau}_1-\tau_0}$

$$\begin{aligned} \mu_1-\hat{\mu}_0 &= \mu_1 - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0 - \frac{(\hat{\tau}_1-\tau_1)\mu_1-(\tau_1-\tau_0)\mu_0}{\hat{\tau}_1-\tau_0} \\ &= \frac{(\hat{\tau}_1-\tau_0)\mu_1-(\hat{\tau}_1-\tau_1)\mu_1-(\tau_1-\tau_0)\mu_0}{\hat{\tau}_1-\tau_0} - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0 \\ &= \frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1-\mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0 \end{aligned}$$

$$\begin{aligned} MISE(\tau_1,\hat{\tau}_1) &= \mathbb{E}\left(\varepsilon_1(\mu_1-\hat{\mu}_0)^2\right) \\ &= \mathbb{E}\left[\varepsilon_1\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1-\mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0\right)^2\right] \\ &\leq \|\varepsilon_1\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1-\mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0\right)^2\right] \end{aligned}$$

$\|\varepsilon_1\|_\infty = sup|\varepsilon_{1_k}| \leq M_{\varepsilon_1}$
$\mathbb{E}\left[\left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}(\mu_1-\mu_0) - \frac{\sigma}{\sqrt{\hat{\tau}_1-\tau_0}}\times U_0\right)^2\right] = \frac{\sigma^2}{\hat{\tau}_1-\tau_0} + \left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2$
Thus :

$$MISE(\tau_1,\hat{\tau}_1) \leq M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1-\tau_0} + \left(\frac{\tau_1-\tau_0}{\hat{\tau}_1-\tau_0}\right)^2(\mu_1-\mu_0)^2\right]$$

$ISE(\hat{\tau}_1,\hat{\tau}_2) = (\hat{\tau}_1,\hat{\tau}_2)(\mu_1-\hat{\mu}_1)^2$
We have :
$\hat{\mu}_1 = \mu_1 + \frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_1}}\times U_1 \Rightarrow \mu_1-\hat{\mu}_1 = -\frac{\sigma}{\sqrt{\hat{\tau}_2-\hat{\tau}_1}}\times U_1$
Thus :

$$\begin{aligned} MISE(\hat{\tau}_1,\tau_2) &= \mathbb{E}\left((\tau_2-\hat{\tau}_1)(\mu_1-\hat{\mu}_1)^2\right) \\ &= (\tau_2-\hat{\tau}_1)\mathbb{E}(\mu_1-\hat{\mu}_1)^2 \\ &= (\tau_2-\hat{\tau}_1)\frac{\sigma^2}{\tau_2-\hat{\tau}_1}\times\mathbb{E}(U_1^2) \\ &= \sigma^2 \end{aligned}$$

$ISE(\tau_2, \hat{\tau}_3) = (\hat{\tau}_3 - \tau_2)(\mu_1 - \hat{\mu_2})^2$
We have :
$\hat{\mu}_2 = \mu_1 + \frac{\sigma}{\sqrt{\hat{\tau}_3 - \tau_2}} \times U_2 \Rightarrow \mu_1 - \hat{\mu_2} = -\frac{\sigma}{\sqrt{\hat{\tau}_2 - \hat{\tau}_1}} \times U_2$
Thus :

$$\begin{aligned}
MISE(\hat{\tau}_1, \hat{\tau}_2) &= \mathbb{E}\left((\hat{\tau}_2 - \hat{\tau}_1)(\mu_1 - \hat{\mu_2})^2\right) \\
&= (\hat{\tau}_2 - \hat{\tau}_1)\mathbb{E}(\mu_1 - \hat{\mu_1}^2) \\
&= (\hat{\tau}_2 - \hat{\tau}_1)\frac{\sigma^2}{\hat{\tau}_2 - \hat{\tau}_1} \times \mathbb{E}(U_2^2) \\
&= \sigma^2
\end{aligned}$$

$ISE(\hat{\tau}_3, \tau_3) = \varepsilon_3(\mu_1 - \hat{\mu}_3)^2$
we have : $\hat{\mu}_3 = \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_3 + \frac{(\tau_3 - \hat{\tau}_3)\mu_1 + (\tau_4 - \tau_3)\mu_3}{\tau_4 - \hat{\tau}_3}$

$$\begin{aligned}
\mu_1 - \hat{\mu}_3 &= \mu_1 - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_3 - \frac{(\tau_3 - \hat{\tau}_3)\mu_1 + (\tau_4 - \tau_3)\mu_3}{\tau_4 - \hat{\tau}_3} \\
&= \frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_3
\end{aligned}$$

$$\begin{aligned}
MISE(\hat{\tau}_3, \tau_3) &= \mathbb{E}\left(\varepsilon_3(\mu_1 - \hat{\mu}_3)^2\right) \\
&= \mathbb{E}\left[\varepsilon_3\left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_3\right)^2\right] \\
&\leq \|\varepsilon_3\|_\infty \times \mathbb{E}\left[\left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_3\right)^2\right]
\end{aligned}$$

$\|\varepsilon_3\|_\infty = sup|\varepsilon_{3_k}| \leq M_{\varepsilon_3}$
$\mathbb{E}\left[\left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}(\mu_1 - \mu_3) - \frac{\sigma}{\sqrt{\tau_4 - \hat{\tau}_3}} \times U_3\right)^2\right] = \frac{\sigma^2}{\tau_4 - \hat{\tau}_3} + \left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}\right)^2 (\mu_1 - \mu_3)^2$
Thus :

$$MISE(\tau_1, \tau_3) \leq M_{\varepsilon_3}\left[\frac{\sigma^2}{\tau_4 - \hat{\tau}_3} + \left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}\right)^2 (\mu_1 - \mu_3)^2\right]$$

Thus, we can deduce that :

$$\begin{aligned}
MISE(\tau_1, \tau_3) &\leq 2\sigma^2 + M_{\varepsilon_1}\left[\frac{\sigma^2}{\hat{\tau}_1 - \tau_0} + \left(\frac{\tau_1 - \tau_0}{\hat{\tau}_1 - \tau_0}\right)^2 (\mu_1 - \mu_0)^2\right] \\
&+ M_{\varepsilon_3}\left[\frac{\sigma^2}{\tau_4 - \hat{\tau}_3} + \left(\frac{\tau_4 - \tau_3}{\tau_4 - \hat{\tau}_3}\right)^2 (\mu_1 - \mu_3)^2\right]
\end{aligned} \tag{B.2}$$

□